Learning Document-Level Semantic Properties from Free-text Annotations

> S.R.K. Branavan, Harr Chen, Jacob Eisenstein, Regina Barzilay MIT

Learning Semantic Properties of Documents

- Commonly formulated in a supervised framework
 - Effective
 - Sensitive to annotation quality and quantity
- Expert annotations are expensive
- Alternative: Use free-text annotations

Free-text annotations



Science: Of Late, Fewer Sunspots Than Usual

Slashdof

Posted by timothy on Tuesday June 10, @12:43PM

esocid writes

"The sun has been laying low for the past couple of years, producing no sunspots and giving a break to <u>satellites</u>. Periods of inactivity are normal for the sun, but this period has gone on longer than usual. The sun usually operates on an 11-year cycle with maximum activity occurring in the middle of the cycle. The last cycle reached its peak in 2001 and is believed to be just ending now, with the next cycle just beginning and expected to reach its peak sometime around 2012. Today's sun, however, is as inactive as it was two years ago, and scientists aren't sure why. In the past, solar physicists observed that the sun once went 50 years without producing sunspots, coinciding with a little ice age on Earth that lasted from 1650 to 1700."

science, space, earth, sol, sun (tagging beta)

Free-text annotations

_	
del.icio.us / tag / photography	
All items tagged photography (create tag description) \rightarrow view popular	Pros: Fabulous Food,Painless Prices,Live Dancing and Singing,Free V Cons: The infrequency of live entertainment
« earlier later »	The Bottom Line: Better than all other Spanish Restaurants by far. Try
Photography Tip: Sharpen Your Digital Photos with the Unsharp Mas by sbombich to photography photoshop lifehacker howto saved by 55 other people	
	kmarie1978's Full Review: Tasca
monoface save this by wests to humor design photography saved by 2003 other people 18 mins ago	Immediately when you walk in, you could forget where you are and mist
Best Photo Sharing Web Site? save this by holden636 to Photography LifeHacks Graphicssaved by 1_other person18 mi	Tapas bar on the streets of Madrid. When there is no live entertainment entertainment, Flamenco dancing, singing and guitar playing, are perfor

Science: Of Late, Fewer Sunspots Than Usual

Slashdof

Posted by timothy on Tuesday June 10, @12:43PM

esocid writes

"The sun has been laying low for the past couple of years, <u>producing no sunspots and giving a break to</u> <u>satellites</u>. Periods of inactivity are normal for the sun, but this period has gone on longer than usual. The sun usually operates on an 11-year cycle with maximum activity occurring in the middle of the cycle. The last cycle reached its peak in 2001 and is believed to be just ending now, with the next cycle just beginning and expected to reach its peak sometime around 2012. Today's sun, however, is as inactive as it was two years ago, and scientists aren't sure why. In the past, solar physicists observed that the sun once went 50 years without producing sunspots, coinciding with a little ice age on Earth that lasted from 1650 to 1700."

science, space, earth, sol, sun (tagging beta)

Free-text annotations



Free-text annotated documents on the web:

- Freely available in large quantities
- Variety of domains

Example: Reviews Pros/Cons

Pros: Fabulous food, Painless prices, Live music, Free valet, free-text labels Cons: The infrequency of live entertainment

kmarie1978's Full Review: Tasca

Immediately when you walk in, you could forget where you are and mistake this beautifully decorate bar area for a Tapas bar on the streets of Madrid. When there is no live entertainment, I prefer to sit upstairs. Live entertainment, Flamenco dancing, singing and guitar playing, are performed downstairs amidst the dining tables. You can be informed of special events at Tasca by joining the email club. Usually a week before an event you will receive an email invitation that you can reply with how many people in your party and how many valet cars.

document

Now getting to the real point, service is mostly Spanish fluent and always friendly. The Sangria is the best I have had in boston, read about it. The Tapas are all I have had at Tasca and I have had every one of them at least once. Here are a list of my favorites in order: Pan con Tomate, Jamon Serrano (curred ham), Queso de Cabra al Horno (roasted tomato and cheese), Albondigas (meatballs in peppercorn sauce), Patatas Bravas (tangy potatoes - mouth watering), and of course the Acietunas (Spanish olives). I could go on and on but I would be copying the menu when you can view it yourself at the website.

Since returning from a trip to Spain, Tasca has most closely replicated the food and atmosphere of Spain Tapas Bars. I prefer this Spanish Tapas restaurant over all the other popular picks in Boston for its fabulous food, unbeatable prices, live entertainment, and great atmosphere for a quiet night with your sweetie or a big get together with friends.

Source: Epinions.com

Free-text Annotations: Label Inconsistency

Pros: *Fabulous food*, Painless prices, Live music, Free valet, *Lovely decor.*

Cons: The infrequency of live entertainment

Pros: Fun place to go with friends; *Excellent food*, Lots of variety, *Beautiful inside*.

Cons: Can get expensive

Free-text Annotations: Missing Labels

Pros: Location, Good beer, Nice place to hang out

Cons: Crowded, Very busy waitstaff

... *The food is way too pricey* for me, a simple suburban girl, so I've only had appetizers. They were good, not plentiful, but tasty and in the five to ten dollar range.

If We had Expert Annotations...

GOOD_FOOD

CHEAP LIVE_MUSIC GOOD_DECOR

kmarie1978's Full Review: Tasca

Immediately when you walk in, you could forget where you are and mistake this beautifully decorate bar area for a Tapas bar on the streets of Madrid. When there is no live entertainment, I prefer to sit upstairs. Live entertainment, I amenco dancing, singing and guitar playing, are performed downstairs amds the dining tables. You can be informed of special events at Tasca by joining the email club. Usually a week before an event you will receive an email invlation that you can reply with how many people in you party and how many valet cars.

Now getting to the real point, service is mostly Spanish fluent and always friendly. The Sangria is the best I have had in boston, read about I. The Tapas are all have had at Tasca and I have had are very one of them at least once. Here are a list of my favorities in order. Pan con Tomate, Jamos Derana (curred harm), Queso de Catar al Homo (roasted tomato and cheese). Albondigas (meatballs in peppercorn sauce). Patatas Bravas (tangy potatoes - mouth watering), and of course the Acietunas (Spanish olives). I could go on and on but I would be copying the menu when you can view it yourself at the website.

Since returning from a trip to Spain, Tasca has most closely replicated the food and atmosphere of Spain Tapas Bars. In prefer this Spanish Tapas restaurant over all the other popular picks in Boston for its fabulous food, unbeatable prices, live entertainment, and great atmosphere for a quiet night with your sweetie or a big get together with friends.

GOOD_FOOD

CHEAP LIVE_MUSIC

vippee1999's Full Review: Tasca

This is a nice "neighborhoody" restaurant with excellent food. As far as I'm concerned, the food is in a whole hother category from Dair of Tapeo (both of which I don't care for). I usually just sit at the bar, so I generally don't have to worry about table availability, but it generally seems to be quite full on any given night of the week, so if you DO want a table, you'd probably be better of making a reservation.

To start, they always bring to your table what tastes like "homemade" bread - very hearty - along with a spread that is usually made of some type of bean spread and/or garlic spread; it's always excellent.

As for dishes, I usually order a few tapas, my favorites being Queso de Cabra (warm goat cheese, on top of broiled tomatoes, served with a basil puree and mixed greens). Confit de Pato (a crissy quick leg, served with a confit of -1 believe - cabbage and apple), and Spanish Sausage in PutP Pastry. I usually follow this up with their dense chocolate torte in a raspberry sauce. Most tapas are in the \$27 range. The cost for three tapas, dessert, and coffee is about \$25.

If you do not have reservations, I would guess that for dinner, one may expect to wait up to 30 minutes, regardless of the day of week. Also, because parking is "tight" in that neighborhood, they do provide complimentary valet parking.

Tasca also does takeout, offering the same food as is on the regular menu. They do a "very nice" takeout, insofar as each of the tapas is individually packed in a styrofoam box. And you still get a good amount of their hearty bread with the fresh spread of the day tog with it.

I really can't say anything "bad" about this restaurant. I eat there about once a month. I should add that the "front" room with the bar has a bit more charm than the "back" room.

Location: 1612/1616? Commonwealth Ave. in Brighton, just past the intersection of Washington Street, in the direction of Cleveland Circle

BAD_FOOD CHEAP --GOOD_DECOR

jgolub's Full Review: Tasca

We been to this restaurant 5 or 6 times now... We gone a couple of times with my family, once on a double date, and a couple of times just with my significant there; in an also a vegetarian, oo ti soften difficult for me to keep going back to the same restaurant, but tasca has several vegetarian options. we usually get 4-5 tapas per couple and taste each others' dishes as well. The presentation is always beautiful; i also like that they bring out dishes a lew at a time, so you never have 20 dishes on the table at the same time... The sample is a sample is a low of tasty, and the desserts are both delicious and pretty, the only drawback i see in this restaurant is that it can get expensive, depending on how hungy you are, which tapas you get, and if you get drinks and dessert. however, it is also an easy place to go and get something small. the wait is never too bad, and there's a small bar to sit at if you are meeting people there.

Identify hidden structure on free-text labels



Learn mapping between text and properties

I am a vegetarian, so it is often difficult for me to keep going back to the same restaurant, but tasca has several vegetarian options. The presentation is always beautiful. the sangria is also very tasty, and the desserts are both delicious and pretty. the only drawback i see in this restaurant is that it can get expensive, depending on how hungry you are, which tapas you get, and if you get drinks and dessert.



tasty preparations great food yummy dishes delicious food

Both tasks performed jointly in a Hierarchical Bayesian Model

Text Authoring as a Generative Process



the sangria is **very tasty**, and the desserts are both **delicious** and pretty. The **tastiest** restaurant food I've had. The only drawback i see it can get expensive, since some of the dishes are **pricey**, and if you get drinks and dessert. You could forget where you are and mistake this **beautifully decorated** bar area for a Tapas bar on the streets of Madrid. It generally seems to be guite **crowded** on any given night of the week.

Text Authoring as a Generative Process



Contrast to Alternative Methods

Topic modelling

e.g. LDA (Blei et al, 2003)



the sangria is **very tasty**, and the desserts are both **delicious** and pretty. The **tastiest** restaurant food I've had. The only drawback i see it can get **expensive**, since some of the dishes are **pricey**, and if you get drinks and dessert. You could forget where you are and mistake this **beautifully decorated** bar area for a Tapas bar on the streets of Madrid. It generally seems to be quite **crowded** on any given night of the week.

Contrast to Alternative Methods

Topic modelling

e.g. LDA (Blei et al, 2003)



the sangria is **very tasty**, and the desserts are both **delicious** and pretty. The **tastiest** restaurant food I've had. The only drawback i see it can get **expensive**, since some of the dishes are **pricey**, and if you get drinks and dessert. You could forget where you are and mistake this **beautifully decorated** bar area for a Tapas bar on the streets of Madrid. It generally seems to be quite **crowded** on any given night of the week.

Supervised Topic Modelling

e.g. sLDA (Blei & McAuliffe, 2007) MAS (Titov & McDonald, 2008)



the sangria is **very tasty**, and the desserts are both **delicious** and pretty. The **tastiest** restaurant food I've had. The only drawback i see it can get **expensive**, since some of the dishes are **pricey**, and if you get drinks and dessert. You could forget where you are and mistake this **beautifully decorated** bar area for a Tapas bar on the streets of Madrid. It generally seems to be quite **crowded** on any given night of the week.

Free-text Labels - Hidden Structure

great price long wait times good price always a queue

This is a nice "neighborhoody" restaurant with excellent food. As far as I'm concerned, the food is in a whole 'nother category from Dali or Tapeo (both of which I don't care for). I usually just sit at the bar, so I generally don't have to worry about table availability, but it generally seems to be quite full on any given night of the week, so if you DO want a table, you'd probably be better off making a reservation.

To start, they always bring to your table what tastes like "homemade" bread - very hearty - along with a spread that is usually made of some type of bean spread and/or garlic spread; it's always excellent.

As for dishes, I usually order a few tapas, my favorites being Queso de Cabra (warm goat cheese, on top of broiled tomatoes, served with a basil puree and mixed greens), Confit de Pato (a crispy duck leg, served with a confit of - I believe - cabbage and apple), and Spanish Sausage in Puff Pastry. I usually follow this up with their dense chocolate torte in a raspberry sauce. Most tapas are in the \$4-7 range. The cost for three tapas, dessert, and coffee is about \$25.

Immediately when you walk in, you could forget where you are and mistake this beautifully decorate bar area for a Tapas bar on the streets of Madrid. When there is no live entertainment, I prefer to sit upstairs. Live entertainment, Flamenco dancing, singing and guitar playing, are performed downstairs amidst the dining tables. You can be informed of special events at Tasca by joining the email club. Usually a week before an event you will receive an email invitation that you can reply with how many people in your party and how many valet cars.

Now getting to the real point, service is mostly Spanish fluent and always friendly. The Sangria is the best I have had in boston, read about it. The Tapas are all I have had at Tasca and I have had every one of them at least once. Here are a list of my favorites in order: Pan con Tomate, Jamon Serrano (curred ham), Queso de Cabra al Horno (roasted tomato and cheese), Albondigas (meatballs in peppercorn sauce), Patatas Bravas (tangy potatoes - mouth watering), and of course the Acietunas (Spanish olives). I could go on and on but I would be copying the menu when you can view it yourself at the website.

Free-text Labels - Hidden Structure

	External similarity information (dictionaries, WordNet, etc)		
great price		good price	
long wait times		always a queue	

This is a nice "neighborhoody" restaurant with excellent food. As far as I'm concerned, the food is in a whole 'nother category from Dali or Tapeo (both of which I don't care for). I usually just sit at the bar, so I generally don't have to worry about table availability, but it generally seems to be quite full on any given night of the week, so if you DO want a table, you'd probably be better off making a reservation.

To start, they always bring to your table what tastes like "homemade" bread - very hearty - along with a spread that is usually made of some type of bean spread and/or garlic spread; it's always excellent.

As for dishes, I usually order a few tapas, my favorites being Queso de Cabra (warm goat cheese, on top of broiled tomatoes, served with a basil puree and mixed greens), Confit de Pato (a crispy duck leg, served with a confit of - I believe - cabbage and apple), and Spanish Sausage in Puff Pastry. I usually follow this up with their dense chocolate torte in a raspberry sauce. Most tapas are in the \$4-7 range. The cost for three tapas, dessert, and coffee is about \$25.

Immediately when you walk in, you could forget where you are and mistake this beautifully decorate bar area for a Tapas bar on the streets of Madrid. When there is no live entertainment, I prefer to sit upstairs. Live entertainment, Flamenco dancing, singing and guitar playing, are performed downstairs amidst the dining tables. You can be informed of special events at Tasca by joining the email club. Usually a week before an event you will receive an email invitation that you can reply with how many people in your party and how many valet cars.

Now getting to the real point, service is mostly Spanish fluent and always friendly. The Sangria is the best I have had in boston, read about it. The Tapas are all I have had at Tasca and I have had every one of them at least once. Here are a list of my favorites in order: Pan con Tomate, Jamon Serrano (curred ham), Queso de Cabra al Horno (roasted tomato and cheese), Albondigas (meatballs in peppercorn sauce), Patatas Bravas (tangy potatoes - mouth watering), and of course the Acietunas (Spanish olives). I could go on and on but I would be copying the menu when you can view it yourself at the website.

Free-text Labels - Hidden Structure



This is a nice "neighborhoody" restaurant with excellent food. As far as I'm concerned, the food is in a whole 'nother category from Dali or Tapeo (both of which I don't care for). I usually just sit at the bar, so I generally don't have to worry about table availability, but it generally seems to be quite full on any given night of the week, so if you DO want a table, you'd probably be better off making a reservation.

To start, they always bring to your table what tastes like "homemade" bread - very hearty - along with a spread that is usually made of some type of bean spread and/or garlic spread; it's always excellent.

As for dishes, I usually order a few tapas, my favorites being Queso de Cabra (warm goat cheese, on top of broiled tomatoes, served with a basil puree and mixed greens), Confit de Pato (a crispy duck leg, served with a confit of - I believe - cabbage and apple), and Spanish Sausage in Puff Pastry. I usually follow this up with their dense chocolate torte in a raspberry sauce. Most tapas are in the \$4-7 range. The cost for three tapas, dessert, and coffee is about \$25.

Immediately when you walk in, you could forget where you are and mistake this beautifully decorate bar area for a Tapas bar on the streets of Madrid. When there is no live entertainment, I prefer to sit upstairs. Live entertainment, Flamenco dancing, singing and guitar playing, are performed downstairs amidst the dining tables. You can be informed of special events at Tasca by joining the email club. Usually a week before an event you will receive an email invitation that you can reply with how many people in your party and how many valet cars.

Now getting to the real point, service is mostly Spanish fluent and always friendly. The Sangria is the best I have had in boston, read about it. The Tapas are all I have had at Tasca and I have had every one of them at least once. Here are a list of my favorites in order: Pan con Tomate, Jamon Serrano (curred ham), Queso de Cabra al Horno (roasted tomato and cheese), Albondigas (meatballs in peppercorn sauce), Patatas Bravas (tangy potatoes - mouth watering), and of course the Acietunas (Spanish olives). I could go on and on but I would be copying the menu when you can view it yourself at the website.

Joint Learning

Use external information



Pairwise label similarity matrix computed from external information

Use label - document relationships





















Parameter estimation using Gibbs sampling

Predicting Document Labels

This is a nice restaurant with excellent food and many delicious options on the menu. They always bring what tastes like "home made" bread - very hearty. On the flipside, it can get quite expensive depending on your order - some of the menu selections are quite pricey. the décor is Spanish and beautiful. It generally seems to be quite full on any given night of the week. If you do not have reservations, I would guess that for dinner, one may expect to wait up to 30 minutes, regardless of the day of week.

Predicting Document Labels

This is a nice restaurant with excellent food and many delicious options on the menu. They always bring what tastes like "home made" bread - very hearty. On the flipside, it can get quite expensive depending on your order - some of the menu selections are quite pricey. the décor is Spanish and beautiful. It generally seems to be quite full on any given night of the week. If you do not have reservations, I would guess that for dinner, one may expect to wait up to 30 minutes, regardless of the day of week.



Predict Supported Topics

Predicting Document Labels

This is a nice restaurant with excellent food and many delicious options on the menu. They always bring what tastes like "home made" bread - very hearty. On the flipside, it can get quite expensive depending on your order - some of the menu selections are quite pricey. the décor is Spanish and beautiful. It generally seems to be quite full on any given night of the week. If you do not have reservations, I would guess that for dinner, one may expect to wait up to 30 minutes, regardless of the day of week.



Data Sets

- Product reviews downloaded from Epinions.com
- Pros / Cons phrases used as free-text annotations

	Restaurants	Cell Phones	Cameras
# of reviews	5735	1112	3971
review length	786.3	1056.9	1014.2
Keyphrases / review	3.42	4.91	4.84

Evaluation

Methods tested on prediction of document labels

Baselines:

- 1. Random baseline
- 2. Label in document
- 3. Label classifier
 - Positive examples : document annotated with label

Model and baselines were compared against:

- Gold standard annotations
- Free-text annotations

Comparison Against Baselines

	Gold Annotation	Free	e-text Annotation	
	Restaurants	Restaurants	Cellphones	Cameras
Random baseline	0.375	0.500	0.494	0.500
Label in text	0.087	0.142	0.259	0.676
Label classifier	0.484	0.666	0.677	0.687
Model	0.517	0.744	0.705	0.743

Our model is significantly better than all baselines at p <= 0.05 (All numbers are F-measures)

Comparison of Joint and Independent Learning

	Gold Annotation	Free	e-text Annotatio	n
	Restaurants	Restaurants	Cellphones	Cameras
Model – Independent Clustering	0.475	0.702 *	0.623 •	0.666 *
Model – Joint Clustering	0.517	0.744	0.705	0.743

Instances against which the joint model performs significantly better are indicated with a * for p <= 0.05 and \bullet for p <= 0.1 (All numbers are F-measures)

Joint learning is better

A System for Summarizing Product Reviews - from Airlines to Zoos



Summary All Reviews
This summary is based on 27 reviews.
Average ratings based on all reviews
Average ratings based on an reviews
Product Rating 🚖 🚖 🚖 🔶 🛛
Bang For The Buck 🛛 🚖 🚖 🚖 🔹
Pros & Cons
+ Brilliant acting
+ Interesting concepts
+ Feel good movie
+ Good character development
Beautiful visuals Bed ending
- Bad ending
- Far too long

Conclusions

- Free-text annotations are a powerful source for learning document semantics
- Our method learns effectively from free-text annotations explicitly handling their incompleteness and inconsistency
- Joint learning of label structure and document label mapping is crucial to this task.